

# Hidden Markov Models for Failure Diagnostic and Prognostic

D. A. Tobon-Mejia<sup>a,b</sup>, K. Medjaher, N. Zerhouni

<sup>a</sup>FEMTO-ST Institute, UMR CNRS  
AS2M Department  
Besançon, France  
diego.tobon@femto-st.fr

G. Tripot

<sup>b</sup>ALSTOM Transport  
R&D Department  
Ornans, France

**Abstract**— This paper deals with an estimation of the Remaining Useful Life of bearings based on the utilization of Mixture of Gaussians Hidden Markov Models (MoG-HMMs). The raw signals provided by the sensors are first processed to extract features, which permit to model the physical component and its degradation. The prognostic process is done in two phases: a learning phase and an evaluation phase. During the first phase, the sensors' data are processed in order to extract appropriate and useful features, which are then used as inputs of dedicated learning algorithms in order to estimate the parameters of a MoG-HMM. The obtained model represents the behavior of the component including its degradation. In addition, the model contains the number of health states and the stay durations in each state. Once the learning phase is done, the generated model is exploited during the second phase, where the extracted features are continuously injected to the learned model to assess the current health state of the physical component and to estimate its remaining useful life and the associated confidence.

The proposed method is tested on a benchmark data taken from the "NASA prognostic data repository" related to bearings used under several operating conditions. Moreover, the developed method is compared to two methods: the first using traditional HMMs with exponential time durations and the second using regular Hidden Semi Markov Model (HSMM). Finally, simulation results are given and discussed at the end of the paper.

**Keywords**- Monitoring, Diagnostic, Prognostic, Remaining Useful Life, Hidden Markov Models

## I. INTRODUCTION

Maintaining the equipment in operating conditions is an industrial, economical and societal requirement. The maintenance can be curative or preventive. In the first case, the maintenance interventions are done only after a fault has been observed or undergone. Thus, this may lead to non desired situations; especially for sensitive industrial systems (one can cite the recent petrochemical explosion in the Gulf of Mexico). To avoid such cases, a systematic maintenance can be implemented. Moreover, the implementation of a condition Based Maintenance (CBM) [1] can lead to significant benefits. Indeed, contrary to a systematic maintenance, a CBM is done according to the real equipment's health condition, estimated or measured through the sensors present on the industrial equipment. In addition, by using appropriate mathematical or modeling tools, the estimated or measured equipment's state

can be projected in order to predict its future health condition and thus take appropriate decisions.

In the framework of CBM, failure prognostic is considered as one of the main processes, as it permits to estimate what is called Remaining Useful Life (RUL) before failure of a given industrial system [2]. Numerous methods and tools can be used to predict the RUL's value. These methods can be classified into three main groups [1]: model-based prognostic, data-driven prognostic and experience-based prognostic. Model-based prognostic [3] deals with the prediction of the RUL of critical physical components by using mathematical or physical models of the degradation phenomenon (crack by fatigue, wear, corrosion, etc.). The data-driven prognostic [3] aims at transforming the data provided by the sensors into relevant models of the degradation's behavior. Finally, the experience-based prognostic [3] exploits the data gathered from the machine during a long period of time to estimate the parameters of traditional reliability laws. Finally, they are used to do extrapolations and projections to compute the RUL.

In this paper, a failure prognostic method based on the use of the data provided by the sensors is presented. This method belongs then to the category of data-driven approach, and uses mainly Mixture of Gaussians Hidden Markov Models (MoG-HMMs) to model the degradation and to estimate the value of RUL before a potential failure. This tool is chosen for its capability to transform the features extracted from the raw signals to relevant behavioral models representing the physical component and its degradation. The proposed failure prognostic method is done in two phases: a learning phase and an evaluation phase. During the first phase, the raw signals coming from the sensors are processed to extract appropriate and useful features. These features are then used to learn the parameters of the MoG-HMMs. In the second phase, the learned models are exploited in order to assess the current condition of the component and to predict its future one. In addition to this latter, a confidence value can be calculated. The RUL and the confidence are important information that may be used in the decision process. For example, they can be used to delay the maintenance interventions or to stop a machine before its future maintenance due to early fault. The developed prognostic method based on the use of MoG-HMMs is finally compared to two other methods: the first uses simple HMMs and the second is based on Hidden Semi-Markov Models (HSMMs).

The paper is organized as follows: after the introduction, some definitions and the existing prognostic approaches are given in section 2. Section 3 presents the proposed method followed by its comparison with two other methods cited previously, with simulation results are given for each method. Finally, section 4 presents the conclusion and some future works.

## II. FAILURE PROGNOSTIC: DEFINITION AND TAXONOMY

Contrary to fault diagnostic, which consists in detecting and isolating the probable cause of the fault [2], [4] and which is done *a posteriori*, i.e. after the occurrence of the fault, failure prognostic aims at anticipating the time of the failure and thus is done *a priori*, as shown in Fig. 1.

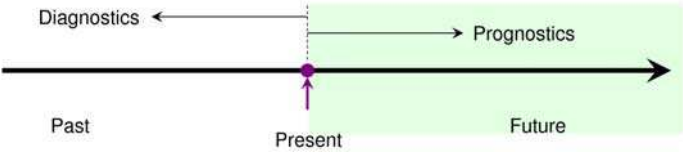


Figure 1. Diagnostic vs prognostic.

Several definitions about failure prognostic have been reported in the recent literature [1], [5], [6]. However, in this paper only that one proposed by the International Standard Organization [7] is considered and where prognostic is defined as “the estimation of the Time To Failure (ETTF) and the risk of existence or later appearance of one or more failure modes”. Note that, most of the definitions reported in the literature use the terminology “Remaining Useful Life (RUL)” instead of “ETTF”. An illustration of a RUL estimation is given in Fig. 2.

In addition to the absolute value of the RUL, a confidence can be associated, which may be useful during the decision process. A method for calculating the confidence value associated to a RUL prediction is proposed in the standard [7]. A list of the factors and the corresponding weights, which may influence the computation of the confidence are suggested.

Numerous methods and tools regarding failure prognostic have been proposed and reported in the literature. This material can be grouped in three main categories [2], [3], [8] (Fig. 3): experience - based prognostic, data - driven prognostic and model - based prognostic.

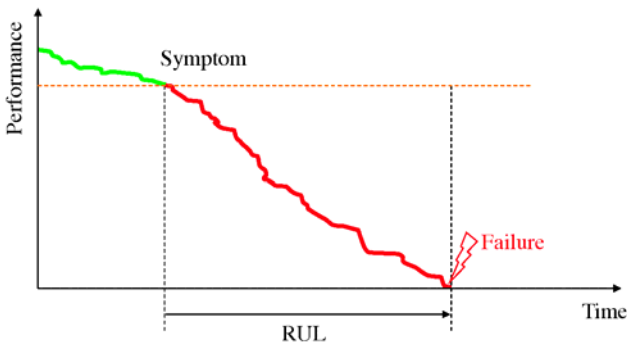


Figure 2. Illustration of a RUL.

1) *Model-based prognostic*: in this approach, the physical component or system and its degradation phenomenon are represented by a set of mathematical laws. The obtained behavioral model is then used to predict the future evolution of the degradation [9], [10]. In this case, the prognostic consists in evolving the degradation model till a determined

future instant, from the actual deterioration state and by considering the future use conditions of the corresponding component.

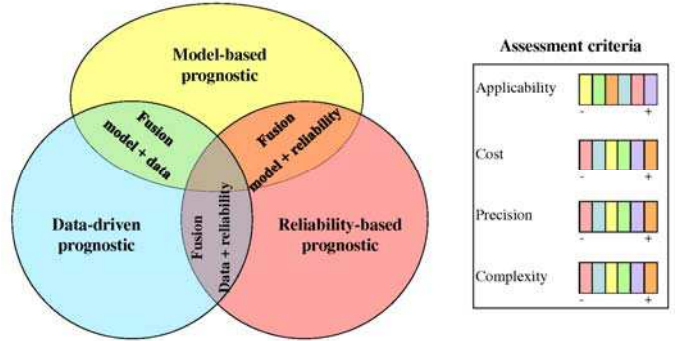


Figure 3. Main prognostic approaches.

The main advantage of this approach dwells in the precision of the obtained results, as the predictions are achieved based on a mathematical model of the degradation. However, the derived degradation model remains specific to a particular kind of component or material, and thus, can not be generalized to all the system components. In addition, getting a mathematical model of the degradation is not an easy task and needs well instrumented test-benches.

2) *Data-driven prognostic*: this approach consists in transforming the monitoring data provided by the sensors installed on the system into reliable behavioral models of the degradations [3], [11]. The collected data are first processed in order to extract relevant features. These latter are then exploited to learn the parameters of behavioral models. The formalization tools used in this approach are mainly those used by the artificial intelligence community, namely: temporal prediction series, trend analysis techniques, neuronal networks under all their facets, neuro-fuzzy systems, hidden Markov models and also dynamic Bayesian networks.

The advantage of this approach is that, for a well monitored system, it is possible to predict the future evolution of the degradation without any need of prior mathematical model of the degradation. However, the results obtained by this approach can be less precise than those of model-based prognostic, and are sometimes considered as local ones (for the case of neural networks and neuro-fuzzy methods).

3) *Experience-based prognostic*: the methods belonging to this approach use probabilistic or stochastic models of the degradation phenomenon, or of the life cycle of the components, by taking into account the data and the knowledge accumulated by experience during the whole exploitation period of the industrial system (experience feedback, maintenance data, etc.). This data is then used to adjust the parameters of some reliability models (Weibull law, exponential law, etc.) [12], [13].

The advantage of this prognostic approach is that it is not necessary to build complex mathematical models. It is also easy to apply on systems for which significant data are available. However, the main drawback of the approach dwells in the amount of data needed to estimate the parameters of the used laws.

### III. HIDDEN MARKOV MODELS

This section gives a brief introduction to the Hidden Markov Models (HMMs) and their variants that will be used in the next section to present the different ways to do failure diagnostic and prognostic in an integrated manner.

An HMM is defined as a statistical model used to represent stochastic processes, where the states are not directly observed [14] (Fig. 4).

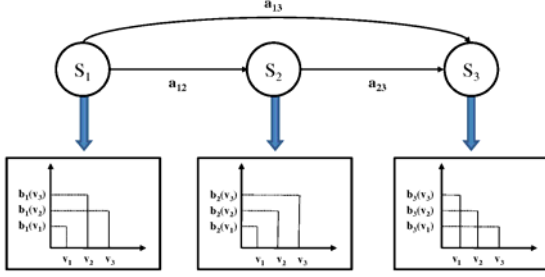


Figure 4. A three state left-to-right HMM.

An HMM is completely defined by the following parameters:

- N: number of states in the model.
- M: the number of distinct observations for each state.
- A: the state transition probability distribution.
- B: the observation probability distribution of each state.
- $\pi$ : the initial state distribution  $\pi$ .

For simplicity and clarity of presentation, a compact notation ( $\lambda = \pi, A, B$ ) is used for each HMM. In practice, HMMs are used to solve three typical problems [14]: a detection problem, a decoding problem and a learning problem.

Usually, discrete HMMs consider the observations as discrete symbols and use discrete probability densities to model the transition and the observation probabilities. The problem with this approach is that in condition monitoring the observations are typically continuous signals. In order to use a continuous observation density, some restrictions are placed to insure that the parameters of the probability density function can be re-estimated [14].

$$b_i(O) = \sum_{m=1}^M C_{jm} \xi(O, \mu_{jm}, U_{jm}), 1 \leq j \leq N \quad (1)$$

In Equation (1), O is the observation vector,  $C_{jm}$  is the mixture coefficient for the  $m^{\text{th}}$  mixture in state  $i$  and  $\xi$  is any log concave or elliptically symmetric density with mean vector  $\mu_{jm}$  and covariance matrix  $U_{jm}$  for the  $m^{\text{th}}$  mixture component in state  $j$ . Usually, a Gaussian density is used for  $\xi$  and the corresponding model is called a MoG-HMM, which is completely defined by: the A matrix, the B matrix and the initial probability  $\pi$ . For a MoG-HMM the observation matrix B is modeled by a Gaussian density with a mean  $\mu$ , a standard deviation  $\sigma$  and a mixture matrix M.

### IV. FAILURE PROGNOSTIC AND DIAGNOSTIC METHODS BASED ON HMMs

Previous research works related to failure diagnostic and prognostic methods based on HMMs and their variants [11], [15] have been developed. The proposed methods are done in two phases: a learning phase and an exploitation phase. In the

following of the paper, an explanation and comparison between three prognostic methods is proposed. The first method is based on traditional HMMs, the second relies on Hidden Semi-Markov Models (HSMM) and the third uses MoG-HMMs.

#### A. The HMMs case

The learning phase and the exploitation phase for use traditional HMMs to perform diagnostic and prognostic are presented below.

1) *Learning phase*: the extracted features from complete monitoring histories (from the normal condition to the failure state) are transformed into HMMs by using the well known Baum-Welch algorithm [16]. Thus, an HMM model is created for each failure, the model is then stored in a model base containing all the HMMs with a diagnostic label associated to each learned histories.

2) *Exploitation phase*: the on-line data are used as inputs to the learned models in order to diagnostic and prognostic the health state. Indeed, the online extracted features are used in a first step to find the model that best fits the actual observation sequence by computing the probability  $P(O|\lambda)$  (similarly to the process in Fig. 5).

In a second step the parameters of the selected model are used to assess the current health state and to estimate the RUL. The Viterbi Algorithm [17] is used to find the state sequence, then the most persistent state in the last observations is retained as the current health state, as shown in (2).

$$\text{State sequence} = (s_1, s_2, \dots, s_t),$$

$$\text{Last states} = (s_{t-l}, \dots, s_{t-2}, s_{t-1}, s_t), \quad (2)$$

with  $l$  = past observations factor and  $t$  = current time

By knowing the actual health state the Chapman-Kolmogorov equation (3) is used to re-estimate the health state after  $n$  iterations [18]. When the predicted probability of being in the last state reaches a predefined limit  $\varepsilon$ , the RUL can be calculated (4).

$$\hat{P}(n) = \pi_i A^n \quad (3)$$

$$RUL = n \Leftrightarrow \hat{P}(s = s_N | n) = \varepsilon \quad (4)$$

#### B. The HSMMs case

The problem with traditional HMMs is that the durational behavior is usually characterized by a geometrically decaying function. This assumption is a source of inaccurate duration modeling because most of real-life applications do not obey this function. Thus, to solve this problem a model with explicit time durations, called Hidden semi Markov Model, has been proposed [11]. The learning and exploitation phase using this model are performed as follows.

1) *Learning phase*: similarly to a traditional HMM, the parameters of the HSMM are defined by using the history data and the Baum-Welch algorithm. For an HSMM the shape is constrained to a left-to-right model. In addition, for each state the stay durations  $D(s_i)$  are learned by using the Viterbi algorithm. The idea is to use the learned parameters ( $\pi, A, B$ ) and the history data to obtain the whole observation sequence. Then, by taking into account the transition instant between the states, the duration  $D$  can be defined.

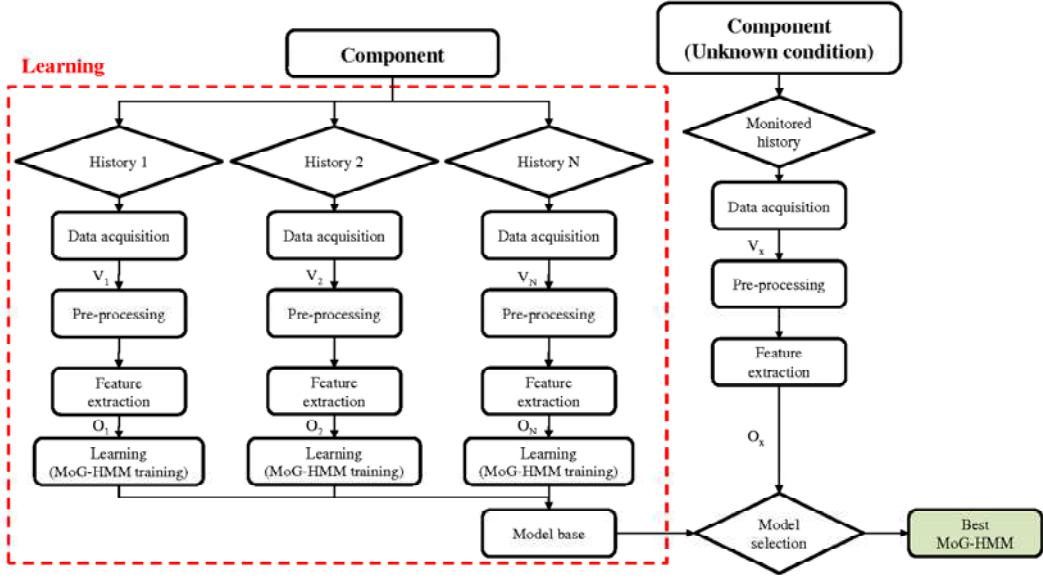


Figure 5. Competitive model selection.

Finally, by assuming that the sojourn time follows a Gaussian distribution, the mean time duration  $\mu(D(s_i))$  and the standard deviation  $\sigma(D(s_i))$  of the same state concerning different histories of the same fault can be estimated ((5) and (6)).

$$\mu(D(s_i)) = \frac{1}{H} \sum_{h=1}^H D(s_i)_h \quad (5)$$

$$\sigma(D(s_i)) = \sqrt{\frac{1}{H} \sum_{h=1}^H [D(s_i)_h - \mu(D(s_i))]^2} \quad (6)$$

In Equations (5) and (6),  $D(\cdot)$  stands for the stay duration,  $i$  is the state index,  $h$  is the history index and  $H$  is the total number of histories from a particular fault state.

2) *Exploitation phase*: in this step the learned models and their associated stay duration are used. First a competitive model selection is performed by using the raw data and the forward-backward algorithm [16] to compute the probability  $P(O|\lambda)$  over all the learned models (Fig. 5). The label of the winner model is used to diagnose the monitored system and the actual health state is defined in the same way as in traditional HMMs using (2). In addition, the selected model is used to estimate the RUL.

The HSMs permits to estimate the RUL in two ways:

a) By using the expression proposed in [11] where the stay durations are merged with the state probability transitions:

$$RUL_l = t_{lc} + a_{l,l+1} [RUL_{l+1}] \quad (7)$$

where  $l$  is the actual state index,  $t_{lc}$  is the state changing point and  $a_{l,l+1}$  is the probability transition to the next state.

b) By using a more intuitive expression adding the state duration from the current state until the last state and subtracting the time spent in the actual state (8).

$$RUL_l = \sum_{i=l}^N \mu(D(s_i)) - t_l \quad (8)$$

Furthermore, a confidence interval with different recovering values can be easily estimated by using the Bonferroni method (9).

$$RUL_{\text{limits}} = \left[ \sum_{i=l}^N \mu(D(s_i)) \pm cf \cdot \sigma(D(s_i)) \right] - t_l \quad (9)$$

$$\Phi(cf) = \frac{\alpha + 1}{2}$$

In Equations (8) and (9),  $l$  is the actual state,  $i$  is the state index,  $N$  the total number of states,  $\mu(D(s_i))$  is the mean time duration in the state  $i$ ,  $\sigma(D(s_i))$  the standard deviation,  $t_l$  is the time spent in the actual state,  $\alpha$  is the confidence interval between  $[0,1]$  and  $cf$  is the confidence factor that are defined by using  $\Phi$  that is the cumulative distribution function of a Gaussian probability distribution.

### C. The MoG-HMM based method

The originality of the proposed method dwells in the fact that raw signals are processed using the Wavelet Packet Decomposition (WPD) to extract the relevant information to learn the behavior models. Also in the generated MoG-HMM, the states' stay durations are not assumed to be a geometrically decaying function as in the HSMs case, but are learned from the monitoring data (note that multiple continuous signals are considered as observation for both learning and simulation phases, instead of the traditional mono observation approach). Moreover, in the proposed method there is no limitation for the type of the generated MoG-HMM (the model can be an ergodic, a left-right or a parallel left-right).

1) *Learning phase*: in this first phase, which is executed off-line, the raw data recorded by the sensors are processed in order to extract the energy of each node at the last decomposition level by using the WPD technique [19]. These features are then used to learn several behavior models (in the form of MoG-HMMs) corresponding to different histories



related to several initial states and/or operating conditions of the component. Also one global left-to-right MoG-HMM is learned for each type of fault. It has  $X$  states meaning the different asset health states. Indeed, each raw data history corresponding to a given component's condition is transformed to a feature matrix  $F$ , by using the WPD. In this matrix, each column vector (of  $C$  features at time  $t$ ) corresponds to a snapshot on the raw signal, and each cell  $f_{ct}$  represents the node  $c$  of the last WPD level at time  $t$ .

$$\text{Raw signal} \xrightarrow{\text{WPD}} F = \begin{pmatrix} f_{11} & \cdots & f_{1t} \\ \vdots & \ddots & \vdots \\ f_{c1} & \cdots & f_{ct} \end{pmatrix} \quad (10)$$

with  $1 \leq t \leq T$  and  $1 \leq c \leq C$

The nodal energy (features) are then used to estimate the parameters ( $\pi$ ,  $A$ ,  $B$ ) and the temporal parameters (stay duration in each state) of the MoG-HMMs.

The parameters of each MoG-HMM are learned using the Baum-Welch algorithm [16], whereas the temporal ones are estimated by using the Viterbi algorithm [17]. This algorithm permits to obtain the state sequence and to compute the time duration for which the component has been in each state of the corresponding MoG-HMM (Fig. 6). Thus, by assuming that the state duration in each state follows a normal law, it is possible to estimate the mean duration (11) and the corresponding standard deviation (12) by computing the duration and the number of visits in each state. Moreover, the Viterbi algorithm permits to identify the final state which represents the physical component's failure state.

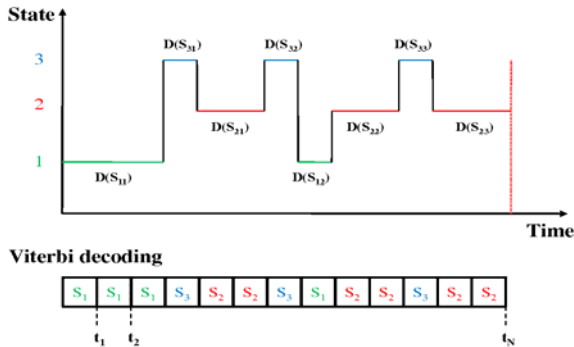


Figure 6. Example of Viterbi decoding state.

$$\mu(D(s_i)) = \frac{1}{\Omega} \sum_{\omega=1}^{\Omega} D(s_{i\omega}) \quad (11)$$

$$\sigma(D(s_i)) = \sqrt{\frac{1}{\Omega} \sum_{\omega=1}^{\Omega} [D(s_{i\omega}) - \mu(D(s_i))]^2} \quad (12)$$

In equations (11) and (12),  $D(\cdot)$  stands for the visit duration,  $i$  is the state index,  $\omega$  is the visit index and  $\Omega$  corresponds to the total of visits. A compact representation of each learned MoG-HMM used to perform diagnostic and prognostic is given by the following expression:

$$\lambda = (\pi, A, B, \mu(D(s_i)), \sigma(D(s_i)), S_{final}) \quad (13)$$

Where  $\lambda$  is the fully defined model and  $S_{final}$  is the final state (corresponding to the end of the considered condition monitoring history).

2) *Exploitation phase*: this phase, which is performed on-line, consists in exploiting the learned models to detect the component's current condition (using the Viterbi algorithm) and to compute the corresponding RUL. The processed data and the extracted nodal energy using the Wavelet toolbox from Matlab® are thus continuously fed to the learned models in order to find the global MoG-HMM, which best fits to the observed sequence, then the diagnostic is made and the current health state is defined. The selection process is based on the calculation of a likelihood  $P(O|\lambda)$  of the model over the observations (HMMs problem 1). Finally, by knowing the current type of fault, the nearest model concerning this fault can be found (Fig. 5). Using this last and the current model state and the stay durations learned in the off-line phase, the component's RUL and its associated confidence value can be estimated.

The generated MoG-HMMs are used during the on-line phase to estimate the RUL and the associated confidence value of the physical component by using a dedicated procedure whose steps are the following:

- Detection of the appropriate global left-to-right general MoG-HMM that best fits and represents the on-line observed sequence of nodal energy. The diagnostic label of the selected model is used to establish the diagnostic of the current condition.
- Choice of the nearest RUL model knowing the active failure mode.
- Identification of the current state of the selected RUL model.
- Identification of the critical path, which goes from the current state to the end state. The idea is to identify all the non-zero probabilities in the transition matrix as potential transitions, and then to choose the minimal path among all the possible ones (Fig. 7) with only one visit per state.

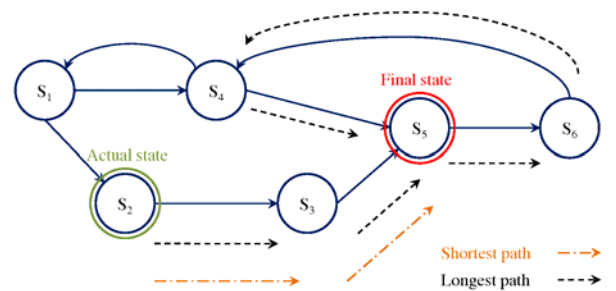


Figure 7. Path estimation

e) Estimation of the RUL. This latter is obtained by using the temporal parameters of the stay duration in each state. In addition, a confidence value over the RUL is calculated based on the standard deviation values of the stay durations and the Bonferroni confidence interval as in (9).

$$RUL_{upper} = \sum_{i=\text{current state}}^N [\mu(D(s_i)) + cf \cdot \sigma(D(s_i))] \quad (14)$$

$$RUL_{mean} = \sum_{i=\text{current state}}^N \mu(D(s_i)) \quad (15)$$

$$RUL_{lower} = \sum_{i=\text{current state}}^N [\mu(D(s_i)) - cf \cdot \sigma(D(s_i))] \quad (16)$$

## V. APPLICATION AND SIMULATION RESULTS

The failure prognostic methods presented previously are tested on a rich condition monitoring data base taken from [20] and containing several bearings tested until the failure. The choice of bearings can be explained by the fact that these components are considered as the most common mechanical elements in industry and are present in almost all industrial processes, especially in those using rotating elements and machines. Moreover, bearings are the main components which most frequently fail in rotating machines [21].

The test data correspond to several tests under constant conditions. Four bearings were installed on one shaft. The angular velocity was kept constant at 2000 rpm and a 6000 lb radial load was applied onto the shaft and bearings (Fig. 8). On each bearing two accelerometers were installed for a total of 8 accelerometers (one vertical Y and one horizontal X) to register the accelerations generated by the vibrations, were the sampling rate was fixed at 20 kHz.

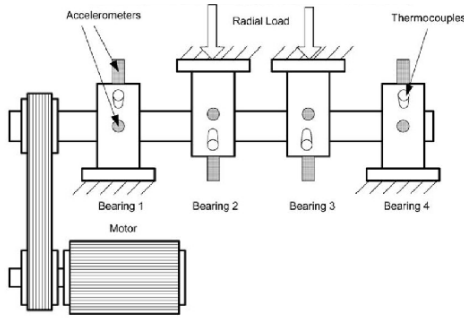


Figure 8. Test bed

For simulation purposes (learning and on-line failure prognostic), twelve condition monitoring data histories are used (eleven for learning and one for test), each bearing was considered failed at the end of its associated history, all the history were supposed independent. In the following results the monitoring history related to the bearing 2 in the test 2 is used as test history. The other eleven monitoring histories are used exclusively for training. The test data history concerns the bearing that fails after 9840 min of operation. Note that for all the models the number of states was fixed to three.

### A. The HMMs case

In traditional HMMs only one discrete observation is used [14]. To transform the continuous signal in a discrete one, the vibrations from the histories are processed. First the root mean square (rms) of the accelerometer placed in the same direction of the load is extracted. Then those values are scaled and rounded to the nearest integer as shown in Fig. 9.

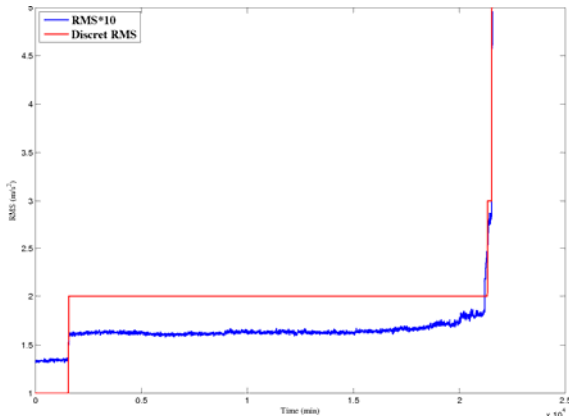


Figure 9. RMS discretisation

The probability transition matrix of the learned model is shown below.

$$A = \begin{pmatrix} 0.9998 & 0.0002 & 0 \\ 0 & 0.9995 & 0.0005 \\ 0 & 0 & 1 \end{pmatrix}$$

By using the Chapman-Kolmogorov equation (3) to find the RUL one can note that the representation of the stay durations by exponential laws in HMMs does not always hold in practice. For example, when the bearing is detected to be in the first health state and using (3) with the learned matrix A, the number of iterations required to get a probability of 95% to be in the failure state (state 3) will be equal to 17518. Knowing that the time interval between the measures is 10 min, this means that the bearing will fail after 175180 min, when the real failure is at 9480 min. The exponential behavior of traditional HMMs are thus a real drawback of this model and that is why other models including the explicit time behavior are wished for prognostic applications.

### B. The HSMMs case

In HSMMs discrete or continuous observations can be used. In this simulation the nodal energies in the third level of the WPD (using the “Daubechies” wavelet family) at each instant t have been extracted from the raw signals (vibration signals). This choice permits to obtain wide frequency bands of 2.5 kHz. The extracted features were used to learn one model per history and then used to predict the RUL at each instant t using (8). The results are presented in Fig. 10, where one can appreciate a dominant late prognostic with an erratic behavior when the system approaches the transition from the state 2 to the state 3, also the model converges late to the real RUL.

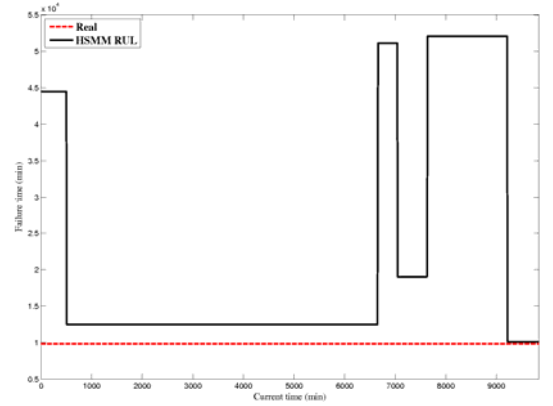


Figure 10. RUL HSMM

### C. The MoG-HMM case

During the learning phase of the MoG-HMM method, the parameters of each model were first randomly initialized and then, the continuous extracted features were fed to the learning algorithms in order to re-estimate the initialized parameters ( $\pi$ , A, B). The number of mixtures in each MoG-HMM was set to two, which allows a trade-off between precision and computation time. Eleven MoG-HMMs were thus obtained by using the Baum-Welch algorithm.

In order to simulate an on-line failure, the “test” data history is used. The selection process shown in Fig. 5 is then applied on the data history corresponding to the bearing 2 in the test 2 in order to identify the “best MoG-HMM” at each instant t and to estimate the RUL. A simulation result of the predicted RUL and the corresponding error is shown in Fig. 11.

From Fig. 12-(a) and Fig. 12-(b) one can see that the precision of the estimated RUL increases as the prediction time is approaching the real failure time. Similarly, after 7440 min the mean estimation error drops below 16.5 % whereas when considering the upper RUL limit the error drops to 7 % and continues to decrease as the real failure time approaches. After 9610 min the mean error stabilizes around the value of 1.08 %. The 68 % confidence interval keeps the RUL estimation limits near acceptable values, a wide confidence intervals (95 % and 99.5 %) will give more sparse limits and will increase the prediction error. Note that the obtained results are better than those obtained by using the temporal features (Root Mean Square, mean, skewness and kurtosis) [22]. The RUL and the associated prediction error of the same test history using temporal features are shown in Fig. 12-(c) and Fig. 12-(d).

#### D. Method Comparison

The results of the precedent methods are compared using the root mean squared error (RMSE) as the model performance criterion. This measure is useful to qualify the accuracy of the predictions thanks to its sensibility to large errors. The RUL estimations of each method were evaluated using the RMSE criterion and the results are presented in Table. 1.

TABLE I. RMSE OF THE PREDICTION ERROR FOR EACH METHOD

Method	HMM	HSMM	WPD/MoG-HMM
RMSE	$1.2322 \times 10^5$	$2.8229 \times 10^4$	$4.6679 \times 10^3$

The results obtained with traditional HMMs are the less performing due to the exponential behavior of the model. The

time integration in HMMs by using explicit state duration like in HSMM improves the prognostic results reducing the error. Finally the proposed MoG-HMM model upgrades the HSMM thanks to its ability to build the more appropriate model unconstraining the model shape and well modeling the continuous signals.

#### VI. CONCLUSIONS

A presentation and comparison of three different methods to perform failure diagnostic and prognostic in an integrated manner has been presented in this paper. Particularly the third method is a new proposition to estimate the current health condition of physical components, particularly bearings, and a prediction of their remaining useful life before their complete failure has been proposed in this paper. The method is based on the transformation of the data provided by the sensors installed to monitor the component into relevant models. These latter are represented by MoG-HMMs, which take as input continuous observations and permit to model the state of the component at each time. The proposed method intends to enhance the existent HSMM to unconstraint the form of the model to be able to model any system. A WPD technique has been used to extract appropriate features from the monitoring signals. These features were then used to model the degradation behavior of the component by learning the parameters of the corresponding MoG-HMMs models. The derived models are finally exploited to asses the component's current condition and to estimate its RUL and the associated confidence value.

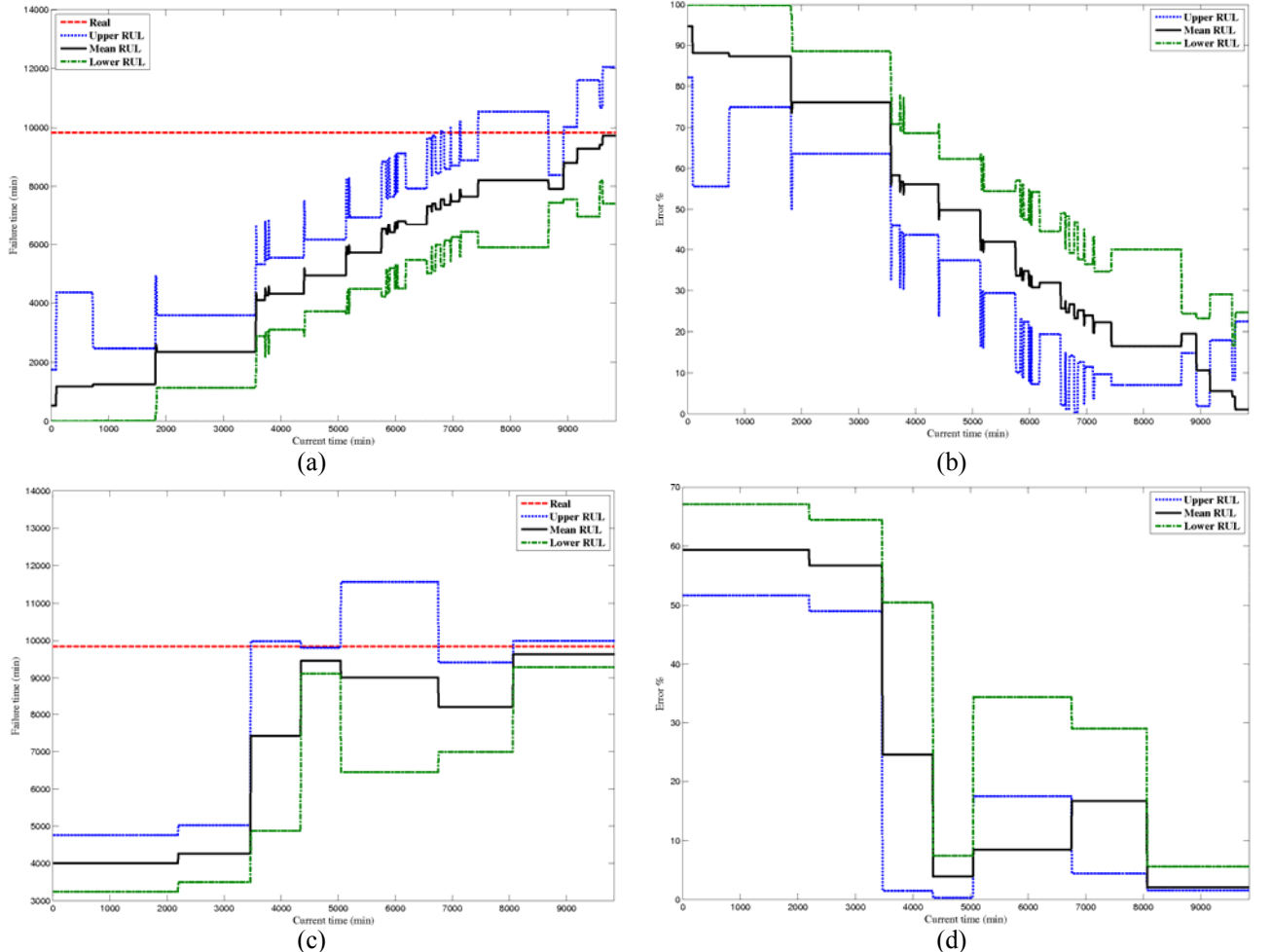


Figure 11. Simulation results: (a) RUL estimation for the bearing 2 in test N°2 using WPD; (b) prediction error for the bearing 2 in test N°2 using WPD; (c) RUL estimation for the bearing 2 in test N°2 using temporal features; and, (d) prediction error for the bearing 2 in test N°2 using temporal features.

## REFERENCES

- [1] M. Lebold and M. Thurston, "Open standards for condition-based maintenance and prognostic systems," in Maintenance and Reliability Conference (MARCON), 2001.
- [2] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483 – 1510, 2006.
- [3] A. Heng, S. Zhang, A. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724 – 739, 2009.
- [4] V. Venkatasubramanian, "Prognostic and diagnostic monitoring of complex systems for product lifecycle management: Challenges and opportunities," *Computers & Chemical Engineering*, vol. 29, no. 6, pp. 1253 – 1263, 2005.
- [5] W. Q. Wang, M. F. Golnaraghi, and F. Ismail, "Prognosis of machine health condition using neuro-fuzzy systems," *Mechanical Systems and Signal Processing*, vol. 18, no. 4, pp. 813 – 831, 2004.
- [6] A. Muller, M.-C. Suhner, and B. Iung, "Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system," *Reliability Engineering & System Safety*, vol. 93, no. 2, pp. 234 – 253, 2008.
- [7] AFNOR, "Condition monitoring and diagnostics of machines - prognostics - part 1: General guidelines. NF ISO 13381-1," 2005.
- [8] G. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent fault diagnosis and prognosis for engineering systems*. Wiley, 2006.
- [9] J. Luo, K. R. Pattipati, L. Qiao, and S. Chigusa, "Model-based prognostic techniques applied to a suspension system," *Transactions on Systems, Man, and Cybernetics*, vol. 38, pp. 1156–1168, 2003.
- [10] D. Chelidze and J. Cusumano, "A dynamical systems approach to failure prognosis," *Journal of Vibration and Acoustics*, vol. 126, pp. 2 – 8, 2004.
- [11] M. Dong and D. He, "A segmental hidden semi-markov model (hsmm)-based diagnostics and prognostics framework and methodology," *Mechanical Systems and Signal Processing*, vol. 21, pp. 2248–2266, 2007.
- [12] A. Keller, U. Perera, and A. Kamath, "Reliability analysis of cnc machine tools," *Reliability Engineering*, vol. 3, no. 6, pp. 449 – 473, 1982.
- [13] P. Groer, "Analysis of time-to-failure with a weibull model," in *Proceedings of the Maintenance and Reliability Conference*, 2000, p. 59.01 - 59.04.
- [14] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, vol. 77 (2), 1989, pp. 257–286.
- [15] H. Ocak, K. A. Loparo, and F. M. Discenzo, "Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics," *Journal of sound and vibration*, vol. 302, pp. 951–961, 2007.
- [16] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, pp. 1 – 38, 1977.
- [17] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm," *IEEE Transaction on Information Theory*, vol. 13, pp. 260 – 269, 1967.
- [18] S. M. Ross, *Probability Models for Computer Science*. Academic Press, 2001.
- [19] J. Zarei and J. Poshtan, "Bearing fault detection using wavelet packet transform of induction motor stator current," *Tribology International*, vol. 40, no. 5, pp. 763 – 769, 2007.
- [20] NSF I/UCRC Center for Intelligent Maintenance Systems, "Prognostic data repository: Bearing data set," in <http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/>, online in 2010.
- [21] P. O'Donnell, "Report of large motor reliability survey of industrial and commercial installations, part I, II & III," *IEEE Transactions on Industry Applications*, vol. 21, pp. 853–872, 1985.
- [22] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "A mixture of gaussians hidden markov model for failure diagnostic and prognostic," in *IEEE Conference on Automation Science and Engineering, CASE'10*, 2010.